TESTIMONY BEFORE THE HOUSE COMMITTEE ON GOVERNMENT REFORM

SUBCOMMITTEE ON TECHNOLOGY, INFORMATION POLICY, INTERGOVERNMENTAL RELATIONS, AND THE CENSUS

BY JEFFREY ROSEN

MARCH 25, 2003

My name is Jeffrey Rosen. I am an associate professor at the George Washington University Law School and legal affairs editor of The New Republic. It is an honor to submit to the Subcommittee on Technology, Information Policy, Intergovernmental Relations, and the Census this prepared testimony on "Data Mining: Current Applications and Future Possibilities," which is adapted from my forthcoming book *The Naked Crowd: Liberty and Security in an Age of Terror* (Random House.)

My thesis is simple: it's possible to design data mining technologies in ways that strike better or worse balances between liberty and security. But there is no guarantee that the executive branch or the technologists, left to their own devices, will demand and provide technologies that strike the balance in a reasonable way. Congress, therefore, has a special responsibility to provide technological and legal oversight of data mining, to ensure that the most invasive searches are focused on the most serious crimes.

For an example of the kind of design choices and oversight I have in mind, I want to focus on the most controversial data mining technology, the so called Total Information Awareness Program designed by the Department of Defense. Although the House and Senate have voted to block funding for TIA in the foreseeable future, its architecture remains a model for other data mining programs that are currently being evaluated by the federal government.

The TIA program is an example of what Roger Clarke has called "mass dataveillance" – that is, the suspicionless surveillance of large groups of people – as opposed to "personal dataveillance," which Clarke defines as the targeted surveillance of individuals who have been identified in advance as suspicious or dangerous. By analyzing financial records, educational records, travel records, and medical records, as well as criminal and other governmental records, TIA proposed to develop technologies that could create risk profiles of millions of Americans citizens and visitors, looking for suspicious patterns of behavior.

In its unregulated form, as Congress recognized, mass dataveillance along the TIA model poses great threats to privacy as well as promising dubious benefits in increased security. When the government engages in mass dataveillance to conduct general searches of millions of citizens without cause to believe that a crime has been committed, the searches arguably raise the same

dangers in the twenty-first century as the general warrants that the Framers of the Fourth Amendment feared in the eighteenth century. Dataveillance, like a general warrant, gives the government essentially unlimited discretion to search through masses of personal information in search of suspicious activity, without specifying in advance the people, places, or things it expects to find. Both general warrants and dataveillance allow fishing expeditions in which the government is trolling for crimes rather than particular criminals, violating the privacy of millions of innocent people in the hope of finding a handful of unknown and unidentified terrorists.

At the same time, mass dataveillance along the TIA model may not be effective in identifying terrorists and picking them out of the crowd. Unlike people who commit credit card fraud – a form of systematic, repetitive and predictable behavior that fits a consistent profile identified by millions of transaction – there is no reason to believe that terrorists in the future will resemble those in the past. There were only 11 hijackers on 9/11, and those who followed them during the following year weren't Saudi Arabians who went to flight school in Florida: they included Richard Reeves, the English citizen who hid a bomb in his shoe, and had a Jamaican father and an English mother. By trying to identify people who look like the 9/11 hijackers, the profiling scheme is looking for a needle in a haystack, but the color and the shape of the needle keep changing. And because the sample of known terrorists is so small, the profiles are bound to be produce a prohibitive number of "false positives" – that is, passengers whom the system wrongly identifies as a likely terrorist. A profiling system that was has a 50% accuracy rate in identifying terrorists would mean that one out of every two passengers would be wrongly singled out for special searches.

Nevertheless, it's possible to design mass dataveillance systems in ways that strike a better balance between privacy and security. The Defense Advanced Research Project Agency, or DARPA, which brought us the Total Information Awareness Program, has also been studying technologies of "selective revelation," which minimize personally identifiable information while allowing data mining and analysis on a grand scale. "The idea of selective revelation is that initially we reveal information to the analyst only in sanitized form, that is, in terms of statistics and categories that do not reveal (directly or indirectly anyone's private information," writes a DARPA report called "Security With Privacy." "If the analyst sees reason for concern he or she can follow up by seeking permission to get more precise information. This permission would be granted if the initial information provides sufficient cause to allow the revelation of more information, under appropriate legal and policy guidelines."[1] But without careful oversight of these secret searches – including audit trails that are reviewed by independent agencies – it's easy to imagine opportunities for abuse.

One way to protect innocent citizens is to ensure that general data searches are constructed in ways that make the data traceable but not easily identifiable – in other words generally anonymous unless officials receive permission to link the data with a particular individual. The Total Information Awareness office has a project called Ginisys that is exploring this kind of architecture for general data searches. According to the program director, Ginisys could protect privacy by separating identifying information from the personal transactions, only recreating the association when there is evidence and legal authority to do so. This might allow, for example, the Center for Disease control to have access to medical information while other groups do not.

The Ginisys staff also plans to develop information privacy filters to keep information that is not relevant out of the repository – encouraging the government to adopt laws that limit the types of data that can be recorded about specific people or transactions. Finally, Ginisys plans to use software agents to mine the information in the repository to expunge information that is unrelated to terrorism.[2]

Although TIA in its current incarnation promises questionable security benefits and grave threats to privacy, the selective revelation architecture poses fewer threats to privacy than the unregulated architecture. How, then, can Congress and the other branches of government encourage the development of mass dataveillance technologies that protect privacy rather than threatening it? Congressional oversight provides the most promising path for America in the twenty-first century. Since the 1960s, Congress has passed more than a dozen important laws protecting privacy, ranging from the Privacy Act of 1974, passed in the wake of Watergate, to the Video Privacy Protection Act of 1988, passed in the wake of journalistic snooping into the video rental records of Robert Bork, the rejected Supreme Court nominee.[3]

As Marc Rotenberg of the Electronic Privacy Information Center has argued, one way for Congress to understand the challenge of balancing liberty and security after 9/11 is the model of checks and balances in the U.S. Constitution. That means that if Congress grants the president new authority engage in foreign intelligence surveillance, it should also create new means of public oversight, or if the Department of Homeland Security proposes a trusted traveler program, it should be subject to open government standards. And if it allows mass dataveillance – whether through the TIA program or the risk profiling systems now being proposed for airports, it should insist on congressional oversight.

Congress could create a special oversight court with the authority to decide when identifying data obtained during mass dataveillance may be connected to transactional information. After intelligence analysts have identified a series of transactions that they believe might be evidence of a terrorist plan, they could petition the special court for authorization to identify the individuals concerned. In considering whether to grant the request, Congress could direct the court to satisfy itself that the crime for which evidence has been presented is a serious threat of force or violence, and that the evidence suggests a links between the suspects and organized terrorists. If the court granted the order, the analysts could link the identifying information with the transaction data, and they could contact federal state and local law enforcement officials to inform them of the threat. In addition to creating this oversight body, and determining legal standards to guide its operation, Congress might also have create standards for federal and citizen oversight, along with penalties for abuse; a dispute resolution process that would give citizens recourse when their data is incorrect or misused; and a series of fair information practices that would give citizens the right to know what personal information the government has collected, and to correct any inaccuracies.

Merely to describe the complexity of these regulations is to raise legitimate questions about whether Congress is ready to adopt them. But Congress has met its oversight responsibilities before. The most important checks on poorly designed technologies of surveillance since 9/11 have come from the Congress – ranging from the decision to block TIA in its current form to the insistence on creating oversight mechanisms for the Carnivore e-mail search program. Rather

than accepting the extreme views of luddites, who believe that all surveillance technology should be resisted, or technopositivists, who believe that no surveillance technology should be regulated, I urge Congress to accept the task of learning about the design choices inherent in these technologies. By evaluating their effectiveness, their necessity, and their impact on privacy Congress can ensure that these technologies are designed in ways that strike reasonable, rather than unreasonable, balances between liberty and security. You have it in your power to strike a thoughtful balance; all you need now is the will.

1."Security with Privacy, ISAT 2002 Study, December 13, 2002, p 10, available at epic.org.

2.Doug Dyer, Address at DAPRATech 2002 Conference, August, 2002.

3.For a comprehensive list of congressional privacy legislation and argument for the superiority of legislative to judicial regulation, see Orin S. Kerr, *The Fourth Amendment in New Technologies: Constitutional Myths and the Case for Restraint* (unpublished draft on file with the author.)